Journal of Multivariate Analysis **75**, 79–111 (2000)
doi:10.1006/jmva.1999.1895, available online at http://www.idealibrary.com on IDEAL®

Shrinkage Estimation towards a Closed Convex Set with a Smooth Boundary

Satoshi Kuriki

The Institute of Statistical Mathematics, Tokyo, Japan
E-mail: kuriki@ism.ac.jp

and

Akimichi Takemura

University of Tokyo, Tokyo, Japan
E-mail: takemura@e.u-tokyo.ac.jp

We give James–Stein type estimators of a multivariate normal mean vector by shrinkage towards a closed convex set K with a smooth or piecewise smooth boundary. The rate of shrinkage is determined by the curvature of the boundary of K at the projection point onto K. By considering a sequence of polytopes K_j converging to K, we show that a particular estimator we propose is the limit of a sequence of shrinkage estimators towards K_j given by M. E. Bock (1982). In fact our estimators reduce to the James–Stein estimator and to the Bock estimator when K is a point and a convex polyhedron, respectively. Therefore they can be considered as natural extensions of these estimators. Furthermore we apply the same method to the problem of improving the restricted mle by shrinkage towards the origin in the multivariate normal mean model where the mean vector is restricted to a closed convex cone with a smooth or piecewise smooth boundary. We demonstrate our estimators in two settings, one shrinking to a ball and the other shrinking to the cone of nonnegative definite matrices.

AMS 1991 subject classifications: 62F10, 60D05, 62F30.

Key words and phrases: James–Stein estimator, second fundamental form, Weyl's tube formula, generalized curvature measure, cone of nonnegative definite matrices.

1. INTRODUCTION

Let x be a p-dimensional random column vector distributed according to the normal distribution $N_p(\mu, I_p)$ with mean vector μ and identity covariance matrix. The problem we consider is estimating the unknown mean vector μ under the loss function

$$L(\hat{\mu}, \mu) = \|\hat{\mu} - \mu\|^2,$$

0047-259X/00 $35.00

the square of the Euclidean norm, and the risk function

$$R(\hat{\mu}, \mu) = E_\mu[L(\hat{\mu}, \mu)].$$

It is very well known that when $p \geqslant 3$ the uniformly minimum variance unbiased estimator x, which is minimax as well, is inadmissible and its risk is improved uniformly in μ by the James–Stein estimator

$$\hat{\mu}(x, \{\mu_0\}) = \mu_0 + \left(1 - \frac{p-2}{\|x-\mu_0\|^2}\right)(x-\mu_0) \tag{1}$$

(James and Stein [7]). Also, let M be a $(p-m)$-dimensional affine subspace in R^p, and denote by x_M the orthogonal projection of x onto M. The estimator

$$\hat{\mu}(x, M) = x_M + \left(1 - \frac{m-2}{\|x-x_M\|^2}\right)(x-x_M) \tag{2}$$

is a version of the James–Stein estimator (1) and dominates the estimator x when $m \geqslant 3$.

Here, (1) and (2) are estimators with x shrinking to the particular point μ_0 and the affine subspace M, and it is reasonable to apply these estimators when $\mu = \mu_0$ and $\mu \in M$ can be considered as *a priori* but vague information on μ.

As an extension of these James–Stein type estimators (1) and (2), Bock [4] considered the case where the unknown mean vector μ is assumed to satisfy several inequalities as *a priori*, vague information. Let K be a closed convex polyhedron which is formed by a set of assumed inequalities and let x_K be the orthogonal projection of x onto K. The estimator proposed by Bock [4] is the estimator with x shrinking in the direction of x_K,

$$\begin{aligned} \hat{\mu}(x, K) &= x_K + \left(1 - \frac{m-2}{\|x-x_K\|^2}\right)(x-x_K) && \text{if} \quad m \geqslant 3, \\ &= x && \text{otherwise}, \end{aligned} \tag{3}$$

where m is the codimension of the face F of K which contains x_K as a relatively interior point, i.e., $p - \dim F = m$. Although (3) is formally the same as (2), it is to be noted that m in (3) is a random variable while m in (2) is a constant. In this paper, we treat the general situation where the convex set K is not necessarily polyhedral. Under some regularity conditions on the surface ∂K of K, we give estimators of μ by shrinkage towards K as natural extensions of (1), (2), and (3). The estimators we focus on are James–Stein type because the relation between the rate of shrinkage and the curvature on the boundary ∂K of K is clearly understood in

James–Stein type estimators. It is feasible that more sophisticated shrinkage estimators lead to further improvements in our setup.

The construction of this paper is as follows. In Section 2 we prepare notations on the geometry of piecewise smooth surfaces of convex sets. Distributions of statistics associated with the orthogonal projection onto K are derived from geometric considerations. Proposed shrinkage estimators towards K dominating x are given in Section 3.1. The rate of shrinkage is shown to be determined by the curvature of ∂K at the projection point of x onto K. Discussions on approximating K by a sequence of polytopes are given in Section 3.2. By considering a sequence of polytopes K_j converging to K, a particular estimator we propose is shown to be the limit of a sequence of estimators $\hat{\mu}(x, K_j)$ in (3). In Section 3.3, we show that our method is applicable to improving the risk of mle in the model where the parameter space is restricted to a closed convex cone. In Section 4, we demonstrate our estimators in two settings. First, we give the shrinkage estimators towards the ball with center μ_0 and radius r. It is shown that when r is sufficiently small the proposed estimators dominate the James–Stein estimator (1). Second, the shrinkage estimators towards the cone of nonnegative definite matrices which is a typical example of a piecewise smooth convex set are considered. The performance of these estimators is investigated by numerical studies.

In related work, Bock [5] gives a different type of shrinkage estimator towards a ball for spherically symmetric distributions; Chang [6], Judge *et al.* [8], and Sengupta and Sen [18] discuss shrinkage estimation when the parameter space is restricted by linear inequalities. For recent developments of shrinkage estimation, see also Robert [14], Rukhin [16], and Kubokawa [10].

2. ORTHOGONAL PROJECTION ONTO A CONVEX SET

In this section we prepare materials from convex analysis and differential geometry, and derive the distributions associated with orthogonal projections onto a closed convex set.

Let K be a closed convex set in R^p. For each $x \in R^p$ the orthogonal projection x_K of x onto K satisfying

$$\|x - x_K\| = \min_{y \in K} \|x - y\|$$

is defined uniquely and we have the unique decomposition

$$x = x_K + (x - x_K). \tag{4}$$

Let ∂K be the boundary of K. For fixed $s \in \partial K$, the normal cone of K at s is defined by

$$N(K, s) = \{ y - s \mid y_K = s \}$$

(Section 2.2 of Schneider [17]). Note that $x - x_K \in N(K, x_K)$. Depending on the dimension of the normal cone $N(K, s)$, the boundary ∂K is decomposed as

$$\partial K = D_1(\partial K) \cup \cdots \cup D_p(\partial K) \tag{5}$$

with

$$D_m(\partial K) = \{ s \in \partial K \mid \dim N(K, s) = m \}.$$

Note that (5) is a disjoint partition of ∂K. Define

$$E_m(\partial K) = \{ x \in R^p \backslash K \mid x_K \in D_m(\partial K) \}.$$

Then we also have a disjoint partition

$$R^p \backslash K = E_1(\partial K) \cup \cdots \cup E_p(\partial K).$$

Here we put a regularity condition on smoothness of ∂K.

Assumption 2.1. $D_m(\partial K)$ is a $(p-m)$-dimensional C^2-manifold consisting of a finite number of relatively open connected components.

Remark 2.1. In this paper we call ∂K "piecewise smooth" if ∂K meets Assumption 2.1. Moreover, we call ∂K "smooth" if ∂K is piecewise smooth and $D_m(\partial K)$, $m \geqslant 2$, are empty.

Fix $x \notin K$ and suppose that $s = x_k \in D_m(\partial K)$. From Assumption 2.1, there exists a C^2 local coordinate system $s = s(\theta)$, $\theta = (\theta^1, ..., \theta^{p-m})$, of $D_m(\partial K)$ in a neighborhood of s. The tangent space $T_{s(\theta)}$ of $D_m(\partial K)$ at $s(\theta)$ is spanned by

$$\left\{ b_a(\theta) = \frac{\partial s}{\partial \theta^a}(\theta),\ a = 1, ..., p - m \right\}.$$

Write an orthonormal basis of $T_{s(\theta)}{}^{\perp}$ as

$$\{ n_\alpha(\theta),\ \alpha = 1, ..., m \}$$

satisfying

$$\langle b_a(\theta), n_\alpha(\theta) \rangle = 0$$

and

$$\langle n_\alpha(\theta), n_\beta(\theta)\rangle = \delta_{\alpha\beta} \qquad \text{(Kronecker's delta)},$$

where $\langle\ ,\ \rangle$ denotes the standard inner product. The metric $G = G(\theta)$ of $D_m(\partial K)$ at $s = s(\theta)$ is

$$G(\theta) = (g_{ab}(\theta))_{1 \leqslant a, b \leqslant p-m}$$

with

$$g_{ab}(\theta) = \langle b_a(\theta), b_b(\theta)\rangle.$$

The (a, b)th element of $G(\theta)^{-1}$ is denoted by $g^{ab}(\theta)$.

Note that $\{b_a(\theta)\}$ are C^1-functions in θ, and that we can choose $\{n_\alpha(\theta)\}$ so as to be of class C^1 as well. For example, $\{n_\alpha(\theta)\}$ obtained by the Gram–Schmidt orthonormalization process of the column vectors of the matrix $I_p - BG^{-1}B'$ are of class C^1 in θ, where $B = (b_1(\theta), ..., b_{p-m}(\theta))$ is a $p \times (p-m)$ matrix. For the differentiability of $\{n_\alpha(\theta)\}$, see also p. 699 of Naiman [13] and the references given there.

The second fundamental form $H_\alpha = H_\alpha(\theta)$ of $D_m(\partial K)$ at $s = s(\theta)$ with respect to the normal direction $n_\alpha(\theta)$ is defined as

$$H_\alpha(\theta) = (h^b_{a\alpha}(\theta))_{1 \leqslant a, b \leqslant p-m}$$

with

$$h^b_{a\alpha}(\theta) = \sum_{c=1}^{p-m} h_{ac\alpha}(\theta)\, g^{cb}(\theta), \qquad h_{ab\alpha}(\theta) = \left\langle -\frac{\partial^2 s}{\partial\theta^a\,\partial\theta^b}(\theta), n_\alpha(\theta)\right\rangle.$$

Since $T_{s(\theta)}{}^\perp$ is the affine hull of $N(K, s)$, we can write an element in $N(K, s)$ as $\sum_{\alpha=1}^m t^\alpha n_\alpha(\theta)$ by introducing a new parameter $t = (t^1, ..., t^m)$. Corresponding to the decomposition (4), we have

$$x = s(\theta) + n(\theta, t) \tag{6}$$

with

$$n(\theta, t) = \sum_{\alpha=1}^m t^\alpha n_\alpha(\theta),$$

which is a local one-to-one transformation of $x \leftrightarrow (\theta, t)$. The Jacobian of the transformation (6) first derived by Weyl [24] is stated in Lemma 2.1 below. Another simpler proof of Lemma 2.1 is given in Appendix A.1.

LEMMA 2.1.

$$dx = \pm |I_{p-m} + H(\theta, t)| \, ds(\theta) \, dt \tag{7}$$

with

$$H(\theta, t) = \sum_{\alpha=1}^{m} t^\alpha H_\alpha(\theta),$$

where

$$ds(\theta) = \sqrt{|G(\theta)|} \, d\theta^1 \cdots d\theta^{p-m}$$

is the volume element of $D_m(\partial K)$, *and* $dx = dx_1 \cdots dx_p$, $dt = dt^1 \cdots dt^m$.

Remark 2.2. Weyl [24] has derived the Jacobian (7) in order to obtain the formula for the volume of tube (Weyl's tube formula).

By means of Lemma 2.1, we can discuss the joint density function of (θ, t) when x is distributed as $N_p(\mu, I_p)$. Note that in our application all eigenvalues of $H(\theta, t)$ are nonnegative, and hence $|I_{p-m} + H(\theta, t)|$ is always positive. The following lemmas hold immediately from Lemma 2.1.

LEMMA 2.2. *Let* $x \sim N_p(\mu, I_p)$. *Then the conditional density of* $t = (t^1, ..., t^m)$ *given* $x_K = s(\theta) \in D_m(\partial K)$ *is*

$$\begin{aligned} f(t \mid \theta) \, dt = {} & e(\theta) \cdot \exp\{ -\tfrac{1}{2} \|n(\theta, t)\|^2 + \langle n(\theta, t), \mu - s(\theta) \rangle \} \\ & \times |I_{p-m} + H(\theta, t)| \, dt \\ & \qquad \text{for } t \text{ such that } n(\theta, t) \in N(K, s(\theta)), \\ = {} & 0 \qquad \text{otherwise}. \end{aligned}$$

Here $e(\theta)$ *is a normalizing constant depending on* θ.

LEMMA 2.3. *Let* $x \sim N_p(\mu, I_p)$. *Denote the length of orthogonal projection by*

$$l = \|x - x_K\| = \|n(\theta, t)\| = \sqrt{\textstyle\sum_\alpha (t^\alpha)^2} \tag{8}$$

and put

$$u = l^{-1} t \in S^{m-1} \qquad (\text{the unit sphere in } R^m). \tag{9}$$

Then the conditional density of l given $x_K = s(\theta) \in D_m(\partial K)$ and u such that $n(\theta, u) \in N(K, s(\theta))$ is

$$\begin{aligned} f(l \mid \theta, u)\, dl &= e(\theta, u) \cdot \exp\{-\tfrac{1}{2}l^2 + l\langle n(\theta, u), \mu - s(\theta)\rangle\} \\ &\qquad \times |I_{p-m} + lH(\theta, u)|\, l^{m-1}\, dl \\ &\qquad\qquad \text{for} \quad l \geqslant 0, \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{10}$$

Here $e(\theta, u)$ is a normalizing constant depending on θ and u.

3. SHRINKAGE ESTIMATION TOWARDS A CONVEX SET

3.1. *Proposed Estimators*

As explained in Section 1, we will discuss estimators with x shrinking in the direction of the orthogonal projection of x onto a closed convex set K,

$$\begin{aligned} \hat{\mu}(x, K) &= x_K + (1 - \phi)(x - x_K) \\ &= s(\theta) + (1 - \phi)\, n(\theta, t), \end{aligned} \tag{11}$$

where

$$\phi = \frac{c(x)}{\|x - x_K\|^2} = \frac{c(\theta, t)}{\|n(\theta, t)\|^2}$$

is the rate of shrinkage. Define $\phi \equiv 0$ when $x \in K$. Note that the estimators (1), (2), and (3) can be written in the form of (11). The problem discussed here is to determine the function $c(x) = c(\theta, t)$ for K with a piecewise smooth boundary.

By applying the Stein method of integration by parts to the conditional density (10), we obtain an unbiased estimator of the risk difference. In Lemma 3.1 and Lemma 3.2 below, we use the symbols l and u which are defined in (8) and (9).

LEMMA 3.1. *Assume that, for each $x \in E_m(\partial K)$, $c(x) = c(\theta, lu)$ is a continuous and piecewise differentiable function in l for fixed (θ, u) and satisfies the boundary condition*

$$\lim_{l \to +0, \infty} \frac{c(\theta, lu)}{l} f(l \mid \theta, u) = 0. \tag{12}$$

Then an unbiased estimator $\widehat{\Delta R}$ *of conditional risk difference between the estimators* $\hat{\mu}(x, K)$ *in* (11) *and the minimax estimator* x *under the conditional distribution* (10), *that is,*

$$E_\mu[\Delta L \mid \theta, u] \qquad \textit{with} \quad \Delta L = \|\hat{\mu}(x, K) - \mu\|^2 - \|x - \mu\|^2,$$

is given by

$$\widehat{\Delta R} = \frac{c^2}{l^2} - 2\frac{1}{l}\frac{\partial c}{\partial l} - 2\frac{c}{l^2}(d-2), \tag{13}$$

where

$$\begin{aligned} d = d(x) = d(\theta, t) &= m + \operatorname{tr} H(\theta, t)(I + H(\theta, t))^{-1} \\ &= m + l \operatorname{tr} H(\theta, u)(I + lH(\theta, u))^{-1}. \end{aligned} \tag{14}$$

Proof. Since

$$\Delta L = \frac{c^2}{l^2} - 2c + 2\frac{c}{l}\langle n(\theta, u), \mu - s(\theta)\rangle,$$

it is sufficient to verify that

$$E_\mu\left[\frac{c}{l}\langle n(\theta, u), \mu - s(\theta)\rangle \mid \theta, u\right] = E_\mu\left[-\frac{1}{l}\frac{\partial c}{\partial l} - \frac{c}{l^2}(d-2) + c \mid \theta, u\right]. \tag{15}$$

Note that (15) is a generalization of Stein's identity for multivariate normal mean vectors (Stein [21]). By virtue of the condition (12), (15) is shown as follows:

$$\begin{aligned} 0 &= \frac{c(\theta, lu)}{l} f(l \mid \theta, u)\bigg|_{+0}^{\infty} \\ &= e(\theta, u)\int_0^\infty \frac{\partial}{\partial l}\exp\{l\langle n(\theta, u), \mu - s(\theta)\rangle\} \\ &\quad \times c(\theta, lu)\exp\{-l^2/2\}\,|I_{p-m} + lH(\theta, u)|\,l^{m-2}\,dl \\ &\quad + e(\theta, u)\int_0^\infty \exp\{l\langle n(\theta, u), \mu - s(\theta)\rangle\} \\ &\quad \times \frac{\partial}{\partial l}[c(\theta, lu)\exp\{-l^2/2\}\,|I_{p-m} + lH(\theta, u)|\,l^{m-2}]\,dl \\ &= \text{lhs of (15)} - \text{rhs of (15)}. \end{aligned}$$

The proof is completed. ∎

Remark 3.1. We will call $d(x)$ in (14) "average codimension" in view of Remark 3.6 below. As a function of l (for fixed θ and u), $d(x) = d(\theta, lu)$ is a nondecreasing function such that

$$\lim_{l \to +0} d(\theta, lu) = m, \qquad \lim_{l \to \infty} d(\theta, lu) = m + \operatorname{rank} H(\theta, u).$$

Lemma 3.1 immediately gives the following.

LEMMA 3.2. *Assume that, for each $x \in E_m(\partial K)$, $c(x) = c(\theta, lu)$ satisfies the conditions of Lemma* 3.1, *and $\widehat{\Delta R} \leqslant 0$ a.e. Then the estimator $\hat{\mu}(x, K)$ in* (11) *is a minimax estimator of μ. Moreover $\hat{\mu}(x, K)$ dominates the estimator x unless $\widehat{\Delta R} = 0$ a.e.*

Now, we give two types of shrinkage estimators towards the convex set K whose boundary is piecewise smooth. One is denoted by $\hat{\mu}(x, K)$ and the other is denoted by $\hat{\mu}^{\dagger}(x, K)$, which are given by Theorems 3.1 and 3.2, respectively.

THEOREM 3.1. *For $x \notin K$, define $\bar{d} = \bar{d}(x) = \bar{d}(\theta, t)$ by*

$$\begin{aligned} \bar{d}(\theta, t) &= d(\theta, t) && \textit{if} \quad m \geqslant 2, \\ &= d(\theta, t) + \frac{1}{|I_{p-1} + H(\theta, t)|} && \textit{if} \quad m = 1, \end{aligned} \tag{16}$$

where $m = \dim N(K, x_K)$. Then, the estimator $\hat{\mu}(x, K)$ in (11) *with $c = \bar{d} - 2$ $(x \notin K)$, $= 0$ $(x \in K)$, is a minimax estimator, and dominates x unless $m + \operatorname{rank} H(\theta, t) \leqslant 2$ a.e. The risk gain $-\Delta R = -E_\mu[\Delta L]$, minus of risk difference, of the estimator is*

$$\begin{aligned} & E_\mu\left[\chi_{\{m \geqslant 2\}} \cdot \frac{1}{l^2}\{(m - 2 + \operatorname{tr} H(I+H)^{-1})^2 + 2 \operatorname{tr} H(I+H)^{-2}\}\right] \\ & + E_\mu\left[\chi_{\{m=1\}} \cdot \frac{1}{l^2}\left\{(\operatorname{tr} H(I+H)^{-1} - 1)^2 + 2 \operatorname{tr} H(I+H)^{-2}\right.\right. \\ & \qquad \left.\left. - \frac{1}{|I+H|^2} - 2 \operatorname{tr} H(I+H)^{-1} \cdot \frac{1}{|I+H|}\right\}\right], \end{aligned} \tag{17}$$

where $H = H(\theta, t) = lH(\theta, u)$ and $\chi_{\{\cdot\}}$ is the indicator function.

Before proceeding to the proof of Theorem 3.1, we give a lemma which is proved in Appendix A.2.

LEMMA 3.3. *Let B be a symmetric matrix such that $O \leqslant B \leqslant I$, that is, both B and $I-B$ are nonnegative definite. Then*

$$(\operatorname{tr} B)^2 + 1 - |I-B|^2 - 2\operatorname{tr}(B^2) - 2\operatorname{tr} B \cdot |I-B| \geqslant 0, \tag{18}$$

and the equality holds iff rank $B \leqslant 1$.

Proof of Theorem 3.1. The function $c(x) = \bar{d}(x) - 2$ is continuous and differentiable in l.

The unbiased estimator of the risk gain $-\widehat{\Delta R}$ when $m \geqslant 2$ is the content of the expectation in the first term in (17), which is nonnegative. Moreover, $-\widehat{\Delta R}$ is positive for $m \geqslant 3$; or for $m = 2$ and rank $H \geqslant 1$. When $m = 1$, $-\widehat{\Delta R}$ reduces to $1/l^2$ times the lhs of (18) with $B = H(I+H)^{-1}$. Note that rank $B =$ rank H. Therefore, $-\widehat{\Delta R}$ is always nonnegative, and is positive iff $m \geqslant 3$; or $m = 2$ and rank $H \geqslant 1$; or $m = 1$ and rank $H \geqslant 2$. These three cases are summarized as $m +$ rank $H \geqslant 3$.

Now, it remains to check the boundary condition (12). Since $c(x)$ is bounded, the boundary condition (12) holds obviously for $m \geqslant 3$. It holds that $\lim_{l \to \infty} (c(\theta, lu)/l)\, f(l \mid \theta, u) = 0$ for $m = 1, 2$. Moreover when $m = 2$,

$$\frac{c}{l} = \frac{d-2}{l} \to \operatorname{tr} H(\theta, u), \quad f(l \mid \theta, u) \to 0, \qquad \text{as} \quad l \to +0,$$

and hence (12) also holds for $m = 2$. When $m = 1$, we have

$$\frac{c}{l} = \frac{\bar{d}-2}{l} \to 0, \quad f(l \mid \theta, u) \to e(\theta, u), \qquad \text{as} \quad l \to +0,$$

and (12) holds for $m = 1$. The proof is completed. ▮

The estimator given in Theorem 3.1 may seem complicated at first glance. However, the motivation for it will be revealed in Section 3.2.

The estimator $\hat{\mu}^{\dagger}(x, K)$ given in Theorem 3.2 below is a modification of $\hat{\mu}(x, K)$ only in the region $E_1(\partial K)$ and has a simpler form.

THEOREM 3.2. *For $x \in K$, define $d(x) \equiv 0$. Let $m = \dim N(K, x_K)$. The estimator in* (11) *with $c(x) = \max\{d(x) - 2, 0\}$, that is,*

$$\begin{aligned} \hat{\mu}^{\dagger}(x, K) &= x_K + \left(1 - \frac{d(x)-2}{\|x - x_K\|^2}\right)(x - x_K) && \textit{if} \quad d(x) > 2, \\ &= x && \textit{otherwise}, \end{aligned} \tag{19}$$

is a minimax estimator, which dominates x *unless* $m + \operatorname{rank} H(\theta, t) \leqslant 2$ *a.e. The risk gain* $-\Delta R$ *of the estimator is*

$$E_\mu\left[\chi_{\{d(x)>2\}} \cdot \frac{1}{l^2}\{(m-2+\operatorname{tr} H(I+H)^{-1})^2 + 2\operatorname{tr} H(I+H)^{-2}\}\right], \qquad (20)$$

where $H = H(\theta, t) = lH(\theta, u)$.

Proof. The function $c(x) = \max\{d(x)-2, 0\}$ is continuous and piecewise differentiable in l.

The unbiased estimator of the risk gain $-\widehat{\Delta R}$, the content of the expectation in (20), is nonnegative, and positive when $d(x) > 2$. It holds that $d(x) > 2$ with a positive probability unless $m + \operatorname{rank} H \leqslant 2$ a.e. (see Remark 3.1).

For the boundary condition (12), we only have to verify the case $m = 1$. Since $c(x) = c(\theta, lu)$ is bounded, we have $\lim_{l\to\infty}(c(\theta, lu)/l)\, f(l \mid \theta, u) = 0$. Since $d = d(\theta, lu)$ is a continuous nondecreasing function in l, and $d - 2 \to -1$ as $l \to +0$, it holds that

$$\frac{c}{l} = \frac{\max\{d-2, 0\}}{l} \to 0 \qquad \text{as} \quad l \to +0.$$

Noting again that $f(l \mid \theta, u) \to e(\theta, u)$ as $l \to +0$, we see (12) holds for $m = 1$. The proof is completed. ∎

Remark 3.2. Because $\max\{d(x), 2\} = d(x) = \bar{d}(x)$ for $m \geqslant 2$, $\hat{\mu}^\dagger(x, K) = \hat{\mu}(x, K)$ holds for $x \in E_m(\partial K)$, $m \geqslant 2$, and for $x \in K$. For $x \in E_1(\partial K)$, $\bar{d}(x) - 2 \geqslant \max\{d(x)-2, 0\}$ (see also (48) and (51) of Appendix A.3) and the shrinkage by $\hat{\mu}^\dagger(x, K)$ of Theorem 3.2 towards K is less or equal to the shrinkage by $\hat{\mu}(x, K)$ of Theorem 3.1.

Remark 3.3. When K is polyhedral, it holds that $H(\theta, t) \equiv 0$ and hence $d(x)$ in (14) becomes m. In this case the estimators $\hat{\mu}(x, K)$ of Theorem 3.1 and $\hat{\mu}^\dagger(x, K)$ of Theorem 3.2 reduce to Bock's estimator (3). They obviously reduce to the James–Stein estimator (1) or (2) if $K = \{\mu_0\}$ or K is an affine subspace M. Therefore the estimators $\hat{\mu}(x, K)$ and $\hat{\mu}^\dagger(x, K)$ can be considered as extensions of the James–Stein estimator for K with a piecewise smooth boundary.

Remark 3.4. Besides the estimators $\hat{\mu}(x, K)$ and $\hat{\mu}^\dagger(x, K)$, we can give estimators which satisfy the assumptions of Lemma 3.2 and reduce to Bock's estimator (3) when K is polyhedral. For example, the estimator defined by $c(x) = \max\{m-2, 0\}$ is such an estimator. However, we will show in Section 3.2 that the estimator $\hat{\mu}(x, K)$ of Theorem 3.1 is the natural extension of Bock's estimator, and in this paper we focus on the estimator $\hat{\mu}(x, K)$ and its modification $\hat{\mu}^{\dagger}(x, K)$.

Remark 3.5. The estimator given by Theorem 3.2 has a Baranchik type extension ([1]). Let $r(l)$, $l \geqslant 0$, be a continuous and piecewise differentiable function such that $0 \leqslant r(l) \leqslant 2$, $r'(l) \geqslant 0$ a.e. Then the estimator in (11) with $c(x) = r(l) \max\{d(x) - 2, 0\}$, $l = \|x - x_K\|$, is a minimax estimator. The proof is parallel to that of Theorem 3.2 and omitted.

3.2. *Approximation by a Sequence of Polytopes*

Here we confirm that the estimator $\hat{\mu}(x, K)$ defined by Theorem 3.1 is the natural generalization of Bock's estimator (3) by approximating K by a sequence of polytopes. Throughout this subsection we assume that K is compact.

The convex hull of a finite number of points is called polytope. For any compact convex set K, there exists a sequence K_j, $j = 1, 2, ...,$ of polytopes which converges to K in the sense of Hausdorff distance

$$\rho(K_1, K_2) = \inf\{\lambda \geqslant 0 \mid K_1 \subset K_2 + \lambda U \text{ and } K_2 \subset K_1 + \lambda U\},$$

where U is the unit ball in R^p (e.g., Corollary 3.1.7 of Webster [23]).

Consider the estimator $\hat{\mu}(x, K_j)$ of (3) with respect to the polytope K_j. Denote m in (3) by $d(x, K_j)$, $\max\{d(x, K_j), 2\}$ by $\bar{d}(x, K_j)$, $d(x)$ in (14) by $d(x, K)$, and $\bar{d}(x)$ in (16) by $\bar{d}(x, K)$.

THEOREM 3.3. *Let K_j, $j = 1, 2, ...,$ be a sequence of polytopes which converge to K in the sense of Hausdorff distance. Let A be a bounded, Borel-measurable set in R^p satisfying*

$$A \subset E_m(\partial K) \qquad \text{and} \qquad \operatorname{dist}(A, K) > 0,$$

where

$$\operatorname{dist}(A, K) = \inf\{\|x - y\| \mid x \in A, \ y \in K\}.$$

Then it holds that

$$\lim_{j \to \infty} \int_A d(x, K_j)\, dx = \int_A d(x, K)\, dx, \tag{21}$$

$$\lim_{j \to \infty} \int_A \bar{d}(x, K_j)\, dx = \int_A \bar{d}(x, K)\, dx, \tag{22}$$

and

$$\lim_{j\to\infty}\int_A \hat{\mu}(x, K_j)\,dx=\int_A \hat{\mu}(x, K)\,dx, \tag{23}$$

where dx *is the Lebesgue measure of* R^p.

Theorem 3.3 states that the estimator $\hat{\mu}(x, K)$ is the limit of a sequence of Bock's estimators $\hat{\mu}(x, K_j)$ in a sense of weak convergence of measures. The proof is given in Appendix A.3.

Remark 3.6. $d(x, K)$ can be interpreted as an average of codimensions in the sense of (24). Abbreviate $E_k(\partial K_j)$ as E_{kj}. From (21), we see for large j that

$$\begin{aligned}\sum_k k\cdot \mathrm{Vol}(A\cap E_{kj}) &= \int_A d(x, K_j)\,dx\\ &\approx \int_A d(x, K)\,dx = d(x^*, K)\cdot \mathrm{Vol}(A),\end{aligned}$$

where x^* is a point in A, and hence

$$\sum_k k\cdot\frac{\mathrm{Vol}(A\cap E_{kj})}{\mathrm{Vol}(A)}\approx d(x^*, K). \tag{24}$$

The lhs of (24) is the average of codimension k with respect to the ratios of the volume of E_{kj} in A.

3.3. *Improving the Mle Restricted to a Closed Convex Cone*

In this subsection, we consider the multivariate normal mean model $N_p(\mu, I_p)$ where the mean vector μ is restricted to a closed convex cone, say C, in R^p. This is a typical model which has been studied extensively in the field of order restricted inference (Barlow *et al.* [2], Robertson *et al.* [15], and Shapiro [19]). When x is observed, the restricted mle of μ is given as the point x_C which attains $\min_{\mu\in C}\|x-\mu\|$. Sengupta and Sen [18] showed that, when C is polyhedral, the risk of restricted mle x_C is improved by shrinkage towards the origin. We now demonstrate that their proposition holds for more general cones which are not necessarily polyhedral.

A class of shrinkage estimators considered here is of the form

$$(1-\phi)\,x_C \qquad \text{with} \quad \phi=c(x)/\|x_C\|^2. \tag{25}$$

Denote the dual cone of C by C^*. Since

$$x=x_C+x_{C^*}, \qquad \langle x_C, x_{C^*}\rangle=0,$$

the difference of losses between the shrinkage estimator (25) and the restricted mle x_C is

$$\begin{aligned}\Delta L &= \|(1-\phi)\, x_C-\mu\|^2-\|x_C-\mu\|^2\\ &= \|x_{C*}+(1-\phi)\, x_C-\mu\|^2-\|x_{C*}+x_C-\mu\|^2\\ &= \|x_{C*}+(1-\phi)\, x_C-\mu\|^2-\|x-\mu\|^2.\end{aligned}$$

Because $x_{C*}+(1-\phi)\, x_C$ is the estimator (11) with $K=C^*$, the coefficient ϕ of the rate of shrinkage can be determined by the method developed in Section 3.1 as long as C^* satisfies Assumption 2.1.

4. EXAMPLES

In this final section we demonstrate our method by giving shrinkage estimators towards smooth and piecewise smooth convex boundaries.

4.1. *Shrinkage Estimators towards a Ball*

Let $K=\{x \mid \|x-\mu_0\| \leqslant r\}$ be the ball in R^p with center μ_0 and radius r. Let x be a p-dimensional random vector distributed according to $N_p(\mu, I_p)$. We consider the estimators $\hat{\mu}(x, K)$ and $\hat{\mu}^{\dagger}(x, K)$ in Theorems 3.1 and 3.2 by shrinkage towards the ball K. Put $\mu_0=0$ for simplicity.

Assume that $x \notin K$. Then $x_K=(r/\|x\|)\, x$ and $m=\dim N(K, x_K)=1$. By choosing a suitable local coordinate system, we have that

$$H=\frac{l}{r}\, I_{p-1} \qquad \text{with} \quad l=\|x\|-r,$$

and that

$$d(x)=1+(p-1)\,\frac{l}{r+l},$$

$$\bar{d}(x)=d(x)+\left(\frac{r}{r+l}\right)^{p-1}.$$

The estimator $\hat{\mu}(x, K)$ in this model is

$$\begin{aligned}\hat{\mu}_r(x) &= \left(1-\frac{p-2-(r/l)(1-(r/\|x\|)^{p-2})}{\|x\|^2}\right) x \qquad &&\text{if} \quad \|x\|>r,\\ &= x &&\text{otherwise.}\end{aligned} \tag{26}$$

Note that

$$d(x)-2>0 \qquad \Leftrightarrow \qquad \|x\|>\frac{p-1}{p-2}\,r.$$

The estimator $\hat{\mu}^{\dagger}(x, K)$ (19) in this model is

$$\begin{aligned} \hat{\mu}_r^{\dagger}(x) &= \left(1-\frac{p-2-r/l}{\|x\|^2}\right)x \qquad \text{if} \quad \|x\|>\frac{p-1}{p-2}\,r, \\ &= x \qquad \text{otherwise.} \end{aligned} \tag{27}$$

The following theorem states the comparison of the risks of estimators $\hat{\mu}_r$ and $\hat{\mu}_r^{\dagger}$ with those of the James–Stein estimator $\hat{\mu}_{JS}$ of (1) with $\mu_0=0$ and the positive-part James–Stein estimator

$$\hat{\mu}_{JS+}(x)=\max\left(1-\frac{p-2}{\|x\|^2},\,0\right)x. \tag{28}$$

A proof together with the expressions for the risk gains of the estimators are given in Appendix A.4.

THEOREM 4.1. *Assume that* $p\geqslant 3$.

(a) *For sufficiently small* $r>0$, *both* $\hat{\mu}_r$ *in* (26) *and* $\hat{\mu}_r^{\dagger}$ *in* (27) *dominate* $\hat{\mu}_{JS}$.

(b) *For any* $r>0$, $\hat{\mu}_{JS+}$ *in* (28) *does not dominate either* $\hat{\mu}_r$ *or* $\hat{\mu}_r^{\dagger}$.

Remark 4.1. Kubokawa [7, Thm. 4.1] gives a wide class of estimators which dominate the James–Stein estimator. We see that the estimators $\hat{\mu}_r$ and $\hat{\mu}_r^{\dagger}$ do not belong to Kubokawa's class, because in (26) and (27)

$$c(x)\sim p-2-\frac{r}{\sqrt{t}} \qquad \text{as} \quad t=\|x\|^2\to\infty,$$

which does not meet the condition (b) of Theorem 4.1 of Kubokawa [9].

Table 4.1 shows the estimated risk gains

$$-\Delta R=E_{\mu}[\,\|x-\mu\|^2]-E_{\mu}[\,\|\hat{\mu}(x)-\mu\|^2] \tag{29}$$

for the estimators $\hat{\mu}_{JS}$, $\hat{\mu}_{JS+}$, $\hat{\mu}_r$, and $\hat{\mu}_r^{\dagger}$ by Monte Carlo method with 1,000,000 replications. In this study, we put the dimension $p=5$ and $\mu=(\mu_1, 0, ..., 0)$, $0\leqslant\mu_1\leqslant 10$. We can confirm that when $r=0.25$ and 0.5 the risk gains of $\hat{\mu}_{JS}$ are smaller than those of $\hat{\mu}_r$ and $\hat{\mu}_r^{\dagger}$ uniformly in the noncentrality parameter $\lambda=\mu_1^2$. We also see that, for any $r>0$, $\hat{\mu}_{JS+}$ does not

TABLE 4.1

Risk Gains of Estimators by Shrinkage to the Ball
(Monte Carlo Simulation with 1,000,000 Replications)

p	μ_1	$\hat{\mu}_{JS}$	$\hat{\mu}_{JS+}$	$\hat{\mu}_{r=0.25}$	$\hat{\mu}_{r=0.5}$	$\hat{\mu}_{r=1}$	$\hat{\mu}_{r=2}$	$\hat{\mu}^{\dagger}_{r=0.25}$	$\hat{\mu}^{\dagger}_{r=0.5}$	$\hat{\mu}^{\dagger}_{r=1}$	$\hat{\mu}^{\dagger}_{r=2}$
5	0.00	3.00	3.60	3.17	3.32	3.05	1.07	3.19	3.35	2.81	0.41
	0.40	2.90	3.49	3.07	3.21	2.98	1.09	3.08	3.24	2.77	0.43
	0.80	2.64	3.17	2.79	2.92	2.77	1.14	2.80	2.95	2.61	0.51
	1.20	2.29	2.72	2.40	2.51	2.46	1.20	2.41	2.54	2.37	0.62
	1.60	1.90	2.22	1.98	2.07	2.09	1.24	1.98	2.09	2.06	0.74
	2.00	1.53	1.75	1.59	1.65	1.72	1.22	1.59	1.67	1.73	0.86
	2.40	1.22	1.35	1.26	1.30	1.38	1.15	1.26	1.31	1.41	0.93
	2.80	0.97	1.04	1.00	1.03	1.10	1.04	1.00	1.03	1.12	0.93
	3.20	0.78	0.81	0.80	0.82	0.87	0.90	0.80	0.82	0.89	0.88
	3.60	0.64	0.65	0.65	0.66	0.70	0.76	0.65	0.66	0.71	0.79
	4.00	0.52	0.53	0.53	0.54	0.57	0.63	0.53	0.54	0.58	0.67
	4.40	0.44	0.44	0.45	0.45	0.47	0.53	0.45	0.45	0.48	0.56
	4.80	0.37	0.37	0.38	0.38	0.40	0.44	0.38	0.38	0.40	0.47
	5.20	0.32	0.32	0.32	0.33	0.34	0.37	0.32	0.33	0.34	0.39
	5.60	0.28	0.28	0.28	0.28	0.29	0.32	0.28	0.28	0.29	0.33
	6.00	0.24	0.24	0.25	0.25	0.26	0.27	0.25	0.25	0.26	0.28
	6.40	0.21	0.21	0.22	0.22	0.22	0.24	0.22	0.22	0.22	0.24
	6.80	0.19	0.19	0.19	0.19	0.20	0.21	0.19	0.19	0.20	0.21
	7.20	0.17	0.17	0.17	0.17	0.18	0.19	0.17	0.17	0.18	0.19
	7.60	0.15	0.15	0.15	0.16	0.16	0.17	0.15	0.16	0.16	0.17
	8.00	0.14	0.14	0.14	0.14	0.14	0.15	0.14	0.14	0.14	0.15
	8.40	0.13	0.13	0.13	0.13	0.13	0.14	0.13	0.13	0.13	0.14
	8.80	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
	9.20	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
	9.60	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	10.00	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09

dominate either $\hat{\mu}_r$ or $\hat{\mu}^{\dagger}_r$. For example, when $\mu_1 = 6.00$ the risk gain of $\hat{\mu}_{JS+}$ is less than those of $\hat{\mu}_r$ and $\hat{\mu}^{\dagger}_r$.

4.2. *Estimation of Nonnegative Definite Mean Matrix*

Let $\mathscr{S}_p$ be the set of $p \times p$ symmetric matrices. We consider $\mathscr{S}_p$ as a metric vector space with the inner product

$$\langle W_1, W_2 \rangle = \operatorname{tr} W_1 W_2 = \sum_i w_{1ii} w_{2ii} + \sum_{i<j} (\sqrt{2}\, w_{1ij})(\sqrt{2}\, w_{2ij})$$

for $W_1 = (w_{1ij})$, $W_2 = (w_{2ij}) \in \mathscr{S}_p$. The norm is denoted by $\|W\| = \sqrt{\operatorname{tr} W^2}$, $W \in \mathscr{S}_p$. Let C in $\mathscr{S}_p$ be the cone formed by $p \times p$ nonnegative definite matrices, i.e.,

$$C = \{ W \in \mathscr{S}_p \mid W \geqslant O \}.$$

Note that C is self-dual, i.e., $C^* = -C$. As mentioned in Takemura and Kuriki [22], C (and hence C^* as well) is a typical example of closed convex set whose boundary is not smooth but piecewise smooth. The partition (5) of the boundary ∂C is

$$\partial C = \bigcup_{i=1}^{p} D_{i(i+1)/2}(\partial C),$$

where

$$D_{i(i+1)/2}(\partial C) = \{ W \in C \mid \operatorname{rank} W = p - i \}.$$

The statistical model considered here is as follows. Let $X = (x_{ij}) \in \mathcal{S}_p$ be a symmetric random matrix whose components are distributed independently as

$$x_{ii} \sim N(\mu_{ii}, 1), \qquad \sqrt{2}\, x_{ij} \sim N(\sqrt{2}\, \mu_{ij}, 1) \qquad (i < j).$$

Then, the joint distribution of X can be written as

$$\frac{1}{2^{p/2} \pi^{p(p+1)/4}} \exp\{ -\tfrac{1}{2} \|X - M\|^2 \} \prod_{i \leqslant j} dx_{ij},$$

where $M = (\mu_{ij})$ is the mean matrix. Furthermore, we assume that M is nonnegative definite. This model arises as the limit of multivariate variance components model when the number of blocks goes to infinity. See Kuriki [11] and the references given there. We discuss here the estimation of M.

Write the spectral decomposition of $X \in \mathcal{S}_p$ as

$$X = Q \Lambda Q' = (Q_1 \quad Q_2) \begin{pmatrix} \Lambda_1 & O \\ O & \Lambda_2 \end{pmatrix} \begin{pmatrix} Q_1' \\ Q_2' \end{pmatrix},$$

where $l_1 \geqslant \cdots \geqslant l_p$ are the eigenvalues of X, r is the integer such that $l_1 \geqslant \cdots \geqslant l_r \geqslant 0 > l_{r+1} \geqslant \cdots \geqslant l_p$, $\Lambda_1 = \operatorname{diag}(l_1, ..., l_r)$, and $\Lambda_2 = \operatorname{diag}(l_{r+1}, ..., l_p)$. Q_1 and Q_2 are $p \times r$ and $p \times (p - r)$ matrices such that $Q = (Q_1 \quad Q_2)$ is orthogonal. The orthogonal projections of X onto C and its dual cone C^* are given as

$$X_C = Q_1 \Lambda_1 Q_1' \qquad \text{and} \qquad X_{C^*} = Q_2 \Lambda_2 Q_2',$$

respectively.

The restricted mle of M under the restriction $M \in C$ is X_C. According to Section 3.3, we can construct the shrinkage estimator

$$(1-\phi)\, X_C \qquad \text{with} \quad \phi = \frac{c(X)}{\operatorname{tr} \Lambda_1^2}, \tag{30}$$

which dominates X_C.

In order to determine the function $c(X)$, we need the explicit form of the second fundamental form of the boundary of $C^* = -C$, which has been derived by Kuriki and Takemura [12].

LEMMA 4.1. *Non-zero eigenvalues of the second fundamental form $H = H(X_{C^*}, X_C)$ of $D_{r(r+1)/2}(\partial C^*)$ at $X_{C^*} = Q_2 \Lambda_2 Q_2'$ with respect to the normal direction $X_C = Q_1 \Lambda_1 Q_1' \in N(C^*, X_{C^*})$ are*

$$\frac{l_i}{-l_j}, \qquad i = 1, ..., r, \quad j = r+1, ..., p. \tag{31}$$

From Lemma 4.1, d in (14) in this model is

$$\begin{aligned} d(X) &= r(r+1)/2 + \operatorname{tr} H(I+H)^{-1} \\ &= \frac{r(r+1)}{2} + \sum_{i=1}^{r} \sum_{j=r+1}^{p} \frac{l_i}{l_i - l_j}. \end{aligned} \tag{32}$$

$\bar{d}$ in (16) is

$$\begin{aligned} \bar{d}(X) &= d(X) && \text{if} \quad r \geqslant 2, \\ &= d(X) + \prod_{j=2}^{p} \left(\frac{-l_j}{l_1 - l_j} \right) && \text{if} \quad r = 1. \end{aligned}$$

The shrinkage estimator of M by the method of Theorem 3.1 is (30) with

$$\begin{aligned} \phi &= \frac{1}{\sum_{i=1}^{r} l_i^2} \{ \bar{d}(X) - 2 \} && \text{if} \quad r \geqslant 1, \\ &= 0 && \text{otherwise.} \end{aligned}$$

The risk gain (17) for this estimator is

$$E_M[\|X_C - M\|^2] - E_M[\|(1-\phi)\, X_C - M\|^2]$$
$$= E_M\Bigg[\chi_{\{r\geqslant 1\}} \cdot \frac{1}{\sum_{i=1}^r l_i^2}\Bigg\{\Bigg(\frac{r(r+1)}{2} - 2 + \sum_{i=1}^r \sum_{j=r+1}^p \frac{l_i}{l_i - l_j}\Bigg)^2$$
$$+ 2\sum_{i=1}^r \sum_{j=r+1}^p \frac{-l_i l_j}{(l_i - l_j)^2}\Bigg\}$$
$$- \chi_{\{r=1\}} \cdot \frac{1}{l_1^2}\Bigg\{\prod_{j=2}^p \Bigg(\frac{-l_j}{l_1 - l_j}\Bigg)^2 + 2\sum_{j=2}^p \frac{l_1}{l_1 - l_j} \cdot \prod_{j=2}^p \Bigg(\frac{-l_j}{l_1 - l_j}\Bigg)\Bigg\}\Bigg]. \tag{33}$$

The shrinkage estimator by Theorem 3.2 is (30) with

$$\phi^\dagger = \frac{1}{\sum_{i=1}^r l_i^2} \max\{d(X) - 2, 0\} \qquad \text{if} \quad r \geqslant 1,$$
$$= 0 \qquad\qquad \text{otherwise.}$$

The risk gain (20) for this estimator is

$$E_M[\|X_C - M\|^2] - E_M[\|(1-\phi^\dagger)\, X_C - M\|^2]$$
$$= E_M\Bigg[\chi_{\{d(X)>2\}} \cdot \frac{1}{\sum_{i=1}^r l_i^2}\Bigg\{\Bigg(\frac{r(r+1)}{2} - 2 + \sum_{i=1}^r \sum_{j=r+1}^p \frac{l_i}{l_i - l_j}\Bigg)^2$$
$$+ 2\sum_{i=1}^r \sum_{j=r+1}^p \frac{-l_i l_j}{(l_i - l_j)^2}\Bigg\}\Bigg]. \tag{34}$$

Table 4.2 is a numerical study for the proposed estimators. We put the size of matrix $p = 2, 4$, and the nonnegative definite mean matrix $M = \mathrm{diag}(m_1, ..., m_p)$, $m_1 \geqslant \cdots \geqslant m_p \geqslant 0$, without loss of generality. Note that the risk of the estimator X is

$$E_M[\|X - M\|^2] = \tfrac{1}{2} p(p+1),$$

which is 3 for $p = 2$ and 10 for $p = 4$. In Table 4.2, the columns labeled "projection," "shrinkage," and "shrinkage†" are the estimated values of

$$E_M[\|X - M\|^2] - E_M[\|X_C - M\|^2]$$
$$= E_M\Bigg[(p-r)(p-r+1) - \sum_{j=r+1}^p l_j^2 + 2\sum_{i=1}^r \sum_{j=r+1}^p \frac{-l_j}{l_i - l_j}\Bigg], \tag{35}$$

(33), and (34), which are the risk gains by projection to C and by shrinkage towards the origin of the two proposed estimators. We evaluated the

TABLE 4.2

Risk Gains by Projection and Shrinkage
(Monte Carlo Simulation with 1,000,000 Replications)

p	$(m_1, ..., m_p)$	λ	projection	shrinkage	shrinkage[†]
2	$(0, 0)$	0	1.503	0.148	0.148
	$(1, 0)$	1	1.212	0.165	0.165
	$(2, 0)$	4	0.896	0.103	0.103
	$(3, 0)$	9	0.753	0.053	0.053
	$(4, 0)$	16	0.688	0.030	0.030
	$(5, 0)$	25	0.651	0.019	0.019
	$(1, 1)/\sqrt{2}$	1	1.134	0.211	0.211
	$(2, 2)/\sqrt{2}$	4	0.524	0.172	0.172
	$(3, 3)/\sqrt{2}$	9	0.149	0.109	0.109
	$(4, 4)/\sqrt{2}$	16	0.026	0.066	0.066
	$(5, 5)/\sqrt{2}$	25	0.003	0.042	0.042
4	$(0, 0, 0, 0)$	0	5.000	3.276	3.267
	$(1, 0, 0, 0)$	1	4.677	3.036	3.036
	$(2, 0, 0, 0)$	4	4.220	2.337	2.338
	$(3, 0, 0, 0)$	9	3.912	1.634	1.635
	$(4, 0, 0, 0)$	16	3.723	1.134	1.134
	$(5, 0, 0, 0)$	25	3.599	0.811	0.811
	$(1, 1, 0, 0)/\sqrt{2}$	1	4.619	3.143	3.143
	$(2, 2, 0, 0)/\sqrt{2}$	4	3.899	2.655	2.655
	$(3, 3, 0, 0)/\sqrt{2}$	9	3.268	2.059	2.059
	$(4, 4, 0, 0)/\sqrt{2}$	16	2.846	1.539	1.539
	$(5, 5, 0, 0)/\sqrt{2}$	25	2.579	1.151	1.151
	$(1, 1, 1, 0)/\sqrt{3}$	1	4.589	3.227	3.227
	$(2, 2, 2, 0)/\sqrt{3}$	4	3.694	2.901	2.901
	$(3, 3, 3, 0)/\sqrt{3}$	9	2.762	2.428	2.428
	$(4, 4, 4, 0)/\sqrt{3}$	16	2.054	1.930	1.930
	$(5, 5, 5, 0)/\sqrt{3}$	25	1.612	1.497	1.497
	$(1, 1, 1, 1)/2$	1	4.572	3.296	3.297
	$(2, 2, 2, 2)/2$	4	3.559	3.106	3.106
	$(3, 3, 3, 3)/2$	9	2.353	2.761	2.761
	$(4, 4, 4, 4)/2$	16	1.278	2.328	2.328
	$(5, 5, 5, 5)/2$	25	0.548	1.876	1.876

risk gains by averaging the unbiased estimators given by the rhs's of (35), (33), and (34) by Monte Carlo simulations with 1,000,000 replications.

From the simulation results we see that the risk gains by projection and shrinkage are nonincreasing in each element of the diagonal matrix M. Moreover, under the condition that $\lambda = \|M\|^2$ is fixed, the risk gain by shrinkage is larger as the variation of $m_1, ..., m_p$ is smaller; while the risk gain by projection is larger as the variation of $m_1, ..., m_p$ is larger. Also, although the risk gain of (34) seems slightly larger than that of (33), the difference is very little. In particular, when $p=2$ these values are exactly the same because rank $H=1$ for $r=1$.

Remark 4.2. We have obtained the unbiased estimator of risk difference (13) with d in (32) through the second fundamental form (31). But in this model, we can also derive it directly by the method of Sheena [20].

APPENDIX

A.1. *Proof of Lemma* 2.1

We introduce here a notational convention for indices; when some indices appear twice, the symbols of summation are abbreviated and terms are summed up with respect to the indices.

From the representation by components of (6), we have

$$x_i = s_i + t^\alpha n_{i\alpha},$$

and therefore

$$dx_i = \left(b_{ia} + t^\alpha \frac{\partial n_{i\alpha}}{\partial \theta^a} \right) d\theta^a + n_{i\alpha}\, dt^\alpha, \tag{36}$$

where

$$\frac{\partial s}{\partial \theta^a} = b_a = (b_{1a}, ..., b_{pa})', \qquad n_\alpha = (n_{1\alpha}, ..., n_{p\alpha})'.$$

By multiplying (36) by b_{ib}, and summing up with respect to $i=1, ..., p$, we have

$$b_{ib}\, dx_i = (g_{ab} + t^\alpha h_{ab\alpha})\, d\theta^a. \tag{37}$$

Here we used the relation

$$b_{ib} \frac{\partial n_{i\alpha}}{\partial \theta^a} = -\frac{\partial b_{ib}}{\partial \theta^a} n_{i\alpha} = h_{ab\alpha}.$$

On the other hand, by multiplying (36) by $n_{i\beta}$, and summing up with respect to $i=1, ..., p$, we have

$$n_{i\beta}\, dx_i = t^\alpha n_{i\beta} \frac{\partial n_{i\alpha}}{\partial \theta^a}\, d\theta^a + dt^\beta. \tag{38}$$

Combining (37) and (38), we have

$$\det(b_1, ..., b_{p-m}, n_1, ..., n_m) \prod dx_i = \det(g_{ab} + t^\alpha h_{ab\alpha}) \prod d\theta^a \prod dt^\alpha.$$

Since

$$\begin{aligned}
&|\det(b_1, ..., b_{p-m}, n_1, ..., n_m)| \\
&\quad = \{\det(b_1, ..., b_{p-m}, n_1, ..., n_m)' \, (b_1, ..., b_{p-m}, n_1, ..., n_m)\}^{1/2} \\
&\quad = \det \begin{pmatrix} (g_{ab}) & O \\ O & I_m \end{pmatrix}^{1/2} = \det(g_{ab})^{1/2},
\end{aligned}$$

it holds that

$$\prod dx_i = \pm \det(\delta_a^b + t^\alpha h_{a\alpha}^b) \det(g_{ab})^{1/2} \prod d\theta^a \prod dt^\alpha.$$

The proof is completed.

A.2. *Proof of Lemma* 3.3

We will prove Lemma 3.3 by mathematical induction with respect to the dimension p of the matrix B. Without loss of generality, we assume that $B = \operatorname{diag}(b_i)_{1 \leqslant i \leqslant p}$ with $0 \leqslant b_i \leqslant 1$.

We see easily that the statement holds when $p = 1$.

Assume that the statement holds for the $(p-1) \times (p-1)$ matrix B_{p-1}. The lhs of (18) with

$$B = B_p = \begin{pmatrix} B_{p-1} & 0 \\ 0 & b_p \end{pmatrix}$$

reduces to the quadratic polynomial in b_p,

$$f(b_p) = c_2 b_p^2 + c_1 b_p + c_0,$$

where

$$c_2 = -(|I - B_{p-1}| - 1)^2 \leqslant 0.$$

Note that $c_2 = 0 \Leftrightarrow B_{p-1} = O$. Moreover c_0 is the lhs of (18) with $B = B_{p-1}$, and therefore we have by the assumption of mathematical induction that $c_0 = f(0) \geqslant 0$ and $f(0) = 0 \Leftrightarrow \operatorname{rank} B_{p-1} \leqslant 1$. On the other hand,

$$f(1) = (\operatorname{tr} B_{p-1})^2 + 2 \operatorname{tr} B_{p-1}(I - B_{p-1}) \geqslant 0, \tag{39}$$

and the equality in (39) holds iff $B_{p-1}=O$. Since $f(\cdot)$ is concave, we have $f(b_p)\geqslant 0$ for $0\leqslant b_p\leqslant 1$. The equality $f(b_p)=0$ holds iff

$$\begin{array}{llll} f(0)=0 & & & \text{if} \quad b_p=0, \\ f(1)=0 & & & \text{if} \quad b_p=1, \\ f(0)=f(1)=0 & \text{and} & c_2=0 & \text{if} \quad 0<b_p<1. \end{array}$$

At least one of these three cases holds iff rank $B_p\leqslant 1$. The proof is completed.

A.3. *Proof of Theorem* 3.3

Proof of (21) *and* (22). Fix a bounded open subset A_0 of R^p satisfying $A_0\subset E_m(\partial K)$ and $\text{dist}(A_0, K)>0$. Let $\mathscr{B}(A_0)$ be the Borel field generated by open sets of the topological subspace A_0 in R^p with the relative topology. (21) and (22) are equivalent to the statement that the measures v_j and $\bar{v}_j$ defined by

$$v_j(A)=\int_A d(x, K_j)\,dx \qquad \text{and} \qquad \bar{v}_j(A)=\int_A \bar{d}(x, K_j)\,dx$$

converge weakly to the measures v and $\bar{v}$ defined by

$$v(A)=\int_A d(x, K)\,dx \qquad \text{and} \qquad \bar{v}(A)=\int_A \bar{d}(x, K)\,dx,$$

respectively, where $A\in\mathscr{B}(A_0)$. To prove these, we only have to show that (21) and (22) hold for any open set $A\subset A_0$ (e.g., Thm. III.1 of Bergström [3]). Before preceding to the proof, we prepare some materials mainly from Sections 4.1–4.2 of Schneider [17].

For $x\notin K$ define

$$l_K(x)=\|x-x_K\|>0$$

and

$$u_K(x)=(x-x_K)/l_K(x)\in S^{p-1}.$$

We consider the triplet $(l_K(x), x_K, u_K(x))$ as a point of $R_+\times R^p\times S^{p-1}$.

Let $\zeta\subset R_+\times R^p\times S^{p-1}$ be a bounded open set. Let

$$\tilde{\zeta}=\{l\in R_+ \mid (l, q, u)\in\zeta\}$$

and for $\rho \in \tilde{\zeta}$ let

$$\eta_\rho = \{(q, u) \in R^p \times S^{p-1} \mid (\rho, q, u) \in \zeta\}. \tag{40}$$

Since x_K is continuous in x (and so are $l_K(x)$ and $u_K(x)$) by Lemma 1.8.9 of Schneider [17],

$$A = \{x \in R^p \mid (l_K(x), x_K, u_K(x)) \in \zeta\} \tag{41}$$

is a bounded open set as well. The volume of A in R^p is given by

$$\mathrm{Vol}(A) = \int_{\tilde{\zeta}} d\rho \, \frac{d}{d\rho} \mu_\rho(K, \eta) \bigg|_{\eta = \eta_\rho},$$

where $\mu_\rho(K, \eta)$ is the volume of the local parallel set

$$\{x \in R^p \mid 0 < l_K(x) \leqslant \rho, (x_K, u_K(x)) \in \eta\}.$$

By virtue of the formula for $\mu_\rho(K, \eta)$ (Thm. 4.2.1 of Schneider [17]), we can write

$$\mathrm{Vol}(A) = \int_{\tilde{\zeta}} d\rho \left\{ \frac{1}{p} \sum_{\alpha=1}^{p} \alpha \rho^{\alpha-1} \binom{p}{\alpha} \Theta_{p-\alpha}(K, \eta_\rho) \right\}, \tag{42}$$

where $\Theta_{p-\alpha}(K, \cdot)$ is the generalized curvature measure.

On the other hand, if $A \subset E_m(\partial K)$,

$$\begin{aligned}
\mathrm{Vol}(A) &= \int_A dx \\
&= \int_{\tilde{\zeta}} \rho^{m-1} \, d\rho \int_{(s(\theta), n(\theta, u)) \in \eta_\rho} |I_{p-m} + \rho H(\theta, u)| \, ds(\theta) \, du \\
&= \int_{\tilde{\zeta}} d\rho \sum_{\alpha=m}^{p} \rho^{\alpha-1} \int_{\eta_\rho} \mathrm{tr}_{\alpha-m} H(\theta, u) \, ds(\theta) \, du,
\end{aligned} \tag{43}$$

where $\mathrm{tr}_k H$ is the kth elementary symmetric function of the eigenvalues of H for $k \geqslant 1$ and $\mathrm{tr}_0 H \equiv 1$. Therefore, comparing (42) and (43), we have in this case

$$\begin{aligned}
\frac{\alpha}{p} \binom{p}{\alpha} \Theta_{p-\alpha}(K, \eta) &= \int_\eta \mathrm{tr}_{\alpha-m} H(\theta, u) \, ds(\theta) \, du \qquad &&\text{for} \quad \alpha \geqslant m, \\
&= 0 &&\text{otherwise}.
\end{aligned} \tag{44}$$

Now, by preparing three lemmas we first prove (21).

Let ζ be a bounded open subset of $R_+ \times R^p \times S^{p-1}$, and define A by (41). Assume that for each ρ, η_ρ in (40) is a continuous set of the measure $\Theta_{p-\alpha}(K, \cdot)$ and that $A \subset E_m(\partial K)$ and $\operatorname{dist}(A, K) > 0$. Let

$$A_j = \{x \in R^p \mid (l_{K_j}(x), x_{K_j}, u_{K_j}(x)) \in \zeta\}.$$

Note that for large j, $l_{K_j}(x) > 0$ and $u_{K_j}(x)$ is well-defined.

LEMMA A.1. *As* $j \to \infty$,

$$\int_{A_j} d(x, K_j)\, dx \to \int_A d(x, K)\, dx. \tag{45}$$

Proof. From the property of weak convergence of the generalized curvature measure (Thm. 4.2.1 of Schneider [17]), it holds that

$$\Theta_{p-\alpha}(K_j, \eta_\rho) \to \Theta_{p-\alpha}(K, \eta_\rho). \tag{46}$$

Abbreviate $E_k(\partial K)$ as E_k. Since

$$\operatorname{Vol}(A_j \cap E_k) = \int_{\tilde{\zeta}} \rho^{k-1} d\rho \times \left(\text{the coefficient of } \rho^{k-1} \text{ in } \frac{d}{d\rho} \mu_\rho(K_j, \eta)\right)\bigg|_{\eta=\eta_\rho},$$

we have

$$\begin{aligned}\int_{A_j} d(x, K_j)\, dx &= \sum_{k=1}^{p} k \operatorname{Vol}(A_j \cap E_k)\\ &= \int_{\tilde{\zeta}} d\rho \sum_{k=1}^{p} \frac{k^2}{p} \binom{p}{k} \rho^{k-1} \Theta_{p-k}(K_j, \eta_\rho).\end{aligned} \tag{47}$$

On the other hand, for $x = s(\theta) + ln(\theta, u)$ we have

$$d(x, K) = m + l \operatorname{tr} H(I + lH)^{-1} = \sum_{\alpha=m}^{p} \alpha \frac{l^{\alpha-m} \operatorname{tr}_{\alpha-m} H}{|I + lH|} \tag{48}$$

with $H = H(\theta, u)$, and hence using (44) we have

$$\begin{aligned}\int_A d(x, K)\, dx &= \int_{\tilde{\zeta}} d\rho \int_{\eta_\rho} \sum_{\alpha=m}^{p} \alpha \frac{\rho^{\alpha-m} \operatorname{tr}_{\alpha-m} H}{|I+\rho H|} \cdot |I+\rho H|\, \rho^{m-1}\, ds(\theta)\, du\\ &= \int_{\tilde{\zeta}} d\rho \sum_{\alpha=m}^{p} \alpha \rho^{\alpha-1} \frac{\alpha}{p} \binom{p}{\alpha} \Theta_{p-\alpha}(K, \eta_\rho).\end{aligned} \tag{49}$$

Comparing (47) and (49), we show (45) by (46). ∎

LEMMA A.2. *As* $j \to \infty$,

$$\int_{A_j} d(x, K_j)\, dx - \int_A d(x, K_j)\, dx \to 0. \tag{50}$$

Proof. Since $x_{K_j} \to x_K$, $l_{K_j}(x) \to l_K(x)$, and $u_{K_j}(x) \to u_K(x)$ (Lemma 1.8.9 of Schneider [17]), we see that $\chi_{A_j}(x) \to \chi_A(x)$ for

$$x \notin N = \{x \mid (l_K(x), x_K, u_K(x)) \in \partial\zeta\}.$$

From the assumption on A that η_ρ in (40) is a continuous set of $\Theta_{p-\alpha}(K, \cdot)$, $\mathrm{Vol}(N) = 0$ by (42) and therefore $\chi_{A_j}(x) \to \chi_A(x)$ a.s. Since we can assume that $\{K_j\}$ and hence $\{A_j\}$ as well are uniformly bounded,

$$|\text{lhs of (50)}| \leqslant p \int |\chi_{A_j}(x) - \chi_A(x)|\, dx \to 0. \quad \blacksquare$$

Finally in order to prove (21) for any open set A it suffices to prove the following lemma.

LEMMA A.3. *For each bounded open set* $A \subset R^p$ *satisfying the assumptions of Theorem* 3.3, *there exists a bounded open set* $\zeta \subset R_+ \times R^p \times S^{p-1}$ *satisfying* (41) *such that* η_ρ *defined by* (40) *is a continuous set of the generalized curvature measure* $\Theta_{p-\alpha}(K, \cdot)$.

Proof. Let U_ε be the open ball in R^p with radius ε, and let the linear hull of $N(K, x_K)$ be denoted by $\mathrm{lin}\, N(K, x_K)$. Then, for sufficient small $\varepsilon > 0$,

$$\zeta = \{(\rho, q, u) \mid \rho \in \tilde{\zeta}, (q, u) \in \eta_\rho\}$$

with

$$\begin{aligned}
\tilde{\zeta} &= \{l_K(x) \in R_+ \mid x \in A\},\\
\eta_\rho &= \{(q, u/\|u\|) \in R^p \times S^{p-1} \mid l_K(x) = \rho, x \in A,\\
&\qquad q - x_K \in \mathrm{lin}\, N(K, x_K) \cap U_\varepsilon, u - u_K(x) \in \mathrm{lin}\, N(K, x_K)^\perp \cap U_\varepsilon\}
\end{aligned}$$

is such a set. $\blacksquare$

The proof of (21) is completed.

For proving (22), replace the equations (47) and (48) with

$$\int_{A_j} \bar{d}(x, K_j)\, dx = \sum_{k=1}^{p} \max(k, 2) \cdot \mathrm{Vol}(A_j \cap E_k)$$

and

$$\bar{d}(x, K) = m + l \operatorname{tr} H(I + lH)^{-1} + \chi_{\{m=1\}} \cdot \frac{1}{|I + lH|}$$
$$= \sum_{\alpha=m}^{p} \max(\alpha, 2) \cdot \frac{l^{\alpha-m} \operatorname{tr}_{\alpha-m} H}{|I + lH|}. \tag{51}$$

We see that the proof of (22) is parallel to that of (21). The proof is completed.

Proof of (23). Let $\hat{\mu}_i(x, K_j)$ and $\hat{\mu}_i(x, K)$ be ith coordinates of $\hat{\mu}(x, K_j)$ and $\hat{\mu}(x, K)$, respectively. Then

$$\int_A \hat{\mu}_i(x, K_j)\, dx - \int_A \hat{\mu}_i(x, K)\, dx$$
$$= -\int_A \left\{ \frac{(x - x_{K_j})_i}{\|x - x_{K_j}\|^2} - \frac{(x - x_K)_i}{\|x - x_K\|^2} \right\} (\bar{d}(x, K_j) - 2)\, dx$$
$$- \int_A \frac{(x - x_K)_i}{\|x - x_K\|^2} \{\bar{\nu}_j(dx) - \bar{\nu}(dx)\}. \tag{52}$$

Since $(x - x_{K_j})_i / \|x - x_{K_j}\|^2 - (x - x_K)_i / \|x - x_K\|^2$ converges to 0 on the compact set cl A (the closure of A) and that $|\bar{d}(x, K_j) - 2| \leqslant p - 2$, the first term of rhs of (52) converges to 0. The second term also converges to 0 because $(x - x_K)_i / \|x - x_K\|^2$ is bounded and the measure $\bar{\nu}_j$ converges weakly to $\bar{\nu}$. The proof is completed.

A.4. *Proof of Theorem* 4.1

Note that the risk gains (29) of the estimators $\hat{\mu}_{JS}$, $\hat{\mu}_{JS+}$, $\hat{\mu}_r$, and $\hat{\mu}_r^\dagger$ are written as

$$-\Delta R_{JS} = E_\mu \left[\frac{(p-2)^2}{\|x\|^2} \right],$$

$$-\Delta R_{JS+} = E_\mu \left[\chi_{\{\|x\|^2 > p-2\}} \frac{(p-2)^2}{\|x\|^2} + \chi_{\{\|x\|^2 \leqslant p-2\}} (2p - \|x\|^2) \right],$$

$$-\Delta R_r = E_\mu \left[\chi_{\{\|x\| > r\}} \left\{ \frac{(p-2)^2}{\|x\|^2} + \frac{r}{\|x\|^3} f_1\left(\frac{r}{\|x\|}\right) \right\} \right]$$

$$-\Delta R_r^\dagger = E_\mu \left[\chi_{\{\|x\| > ((p-1)/(p-2))\, r\}} \left\{ \frac{(p-2)^2}{\|x\|^2} + \frac{r}{\|x\|^3} f_2\left(\frac{r}{\|x\|}\right) \right\} \right],$$

respectively, where

$$f_1(t) = \frac{1}{(1-t)^2}\{2 - t - t^{2p-3} - (2p-2)(1-t)\, t^{p-2}\},$$

$$f_2(t) = \frac{2-t}{(1-t)^2}.$$

Proof of (a). The risk differences between $\hat{\mu}_{JS}$ and $\hat{\mu}_r$, and between $\hat{\mu}_{JS}$ and $\hat{\mu}_r^\dagger$ are

$$\Delta R_{JS} - \Delta R_r = E_\mu[\, g_1(\|x\|) - h_1(\|x\|)]$$

and

$$\Delta R_{JS} - \Delta R_r^\dagger = E_\mu[\, g_2(\|x\|) - h_2(\|x\|)],$$

respectively, where

$$g_i(\|x\|) = \chi_{\{\|x\| > c_i\}} \frac{r}{\|x\|^3} f_i\left(\frac{r}{\|x\|}\right),$$

$$h_i(\|x\|) = \chi_{\{\|x\| \leqslant c_i\}} \frac{(p-2)^2}{\|x\|^2}, \qquad i = 1, 2,$$

with

$$c_1 = r, \qquad c_2 = \frac{p-1}{p-2}\, r.$$

Since $f_1(t)$ is a polynomial of $(2p-5)$th degree in t, and

$$\begin{aligned} f_1(t) &= \frac{1}{1-t}\left\{1 + \sum_{j=0}^{2p-4} t^j - (2p-2)\, t^{p-2}\right\} \\ &> \frac{1}{1-t}\{1 + (2p-3)\, t^{\sum_{j=0}^{2p-4} j/(2p-3)} - (2p-2)\, t^{p-2}\} \\ &= \frac{1-t^{p-2}}{1-t} > 1 \qquad \text{for} \quad 0 < t < 1, \end{aligned} \tag{53}$$

we have $f_1(t) \geqslant 1$ for $0 \leqslant t \leqslant 1$. Here the first inequality in (53) holds by the convexity of t^x. Also, $f_2(t) \geqslant 1$ for $0 \leqslant t < 1$ holds obviously. Hence, we have

$$g_i(\|x\|) \geqslant \chi_{\{\|x\| > c_i\}} \frac{r}{\|x\|^3}, \qquad i = 1, 2.$$

Now we evaluate the expectations of h_i and g_i in turn, and then give a lower bound of $E_\mu[g_i - h_i]$.

Since $\|x\|^2$ is distributed according to the noncentral chi-squared distribution with p degrees of freedom and the noncentrality parameter $\lambda = \|\mu\|^2$, it holds that

$$\begin{aligned}
E_\mu[h_i] &= (p-2)^2 \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} e^{-\lambda/2} \int_0^{c_i^2} \frac{1}{v} \cdot \frac{v^{p/2+k-1}}{2^{p/2+k}\Gamma(p/2+k)} e^{-v/2}\, dv \\
&\leqslant (p-2)^2 \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} e^{-\lambda/2} \int_0^{c_i^2} \frac{v^{p/2+k-2}}{2^{p/2+k}\Gamma(p/2+k)}\, dv \\
&= (p-2)^2 \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} e^{-\lambda/2} \frac{c_i^{p+2k-2}}{2^{p/2+k}(p/2+k-1)\,\Gamma(p/2+k)}.
\end{aligned}$$

By putting

$$F(c) = \max_{k \geqslant 0} \frac{c^{2k}}{2^{p/2}(p/2+k-1)\,\Gamma(p/2+k)} < \infty,$$

we have

$$\begin{aligned}
E_\mu[h_i] &\leqslant (p-2)^2\, c_i^{p-2} F(c_i) \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} \cdot \frac{1}{2^k} e^{-\lambda/2} \\
&= (p-2)^2\, c_i^{p-2} F(c_i)\, e^{-\lambda/4}.
\end{aligned} \tag{54}$$

In the case $p \geqslant 4$, it holds that

$$\begin{aligned}
E_\mu[g_i] &\geqslant E_\mu\left[\chi_{\{\|x\| > c_i\}} \frac{r}{\|x\|^3}\right] \\
&= r \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} e^{-\lambda/2} \int_{c_i^2}^{\infty} \frac{1}{v^{3/2}} \cdot \frac{v^{p/2+k-1}}{2^{p/2+k}\Gamma(p/2+k)} e^{-v/2}\, dv \\
&= r \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} e^{-\lambda/2} a_k \bar{G}_{p-3+2k}(c_i^2),
\end{aligned}$$

where

$$a_k = \frac{\Gamma((p-3)/2+k)}{2^{3/2}\Gamma(p/2+k)},$$

and $\bar{G}_\nu(t) = P(\chi^2_\nu \geqslant t)$ is the upper probability of chi-squared distribution with ν degrees of freedom. Noting that

$$a_k \geqslant \tilde{a}_k = \frac{1}{2^{3/2}((p-1)/2+k)((p-3)/2+k)}, \qquad \bar{G}_{p-3+2k}(c_i^2) \geqslant \bar{G}_1(c_i^2),$$

for $p \geqslant 4$, $k \geqslant 0$, and putting

$$M = \min_{k \geqslant 0} (3/2)^k \, \tilde{a}_k < \infty,$$

we have

$$E_\mu[g_i] \geqslant Mr\bar{G}_1(c_i^2) \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!} \cdot \frac{1}{(3/2)^k} e^{-\lambda/2} = Mr\bar{G}_1(c_i^2)\, e^{-\lambda/6}. \quad (55)$$

Therefore, by (54) and (55), we have

$$E_\mu[g_i - h_i] \geqslant Mr\bar{G}_1(c_i^2)\, e^{-\lambda/6} - (p-2)^2\, c_i^{p-2} F(c_i)\, e^{-\lambda/4}. \quad (56)$$

Note that $c_i = r$ or $((p-1)/(p-2))\, r$, and that $\bar{G}_1(\cdot)$ and $F(\cdot)$ are nonincreasing and nondecreasing, respectively. For a sufficiently small $r > 0$, the rhs of (56) is positive for all λ, and the proof in the case $p \geqslant 4$ is completed.

In the case $p = 3$,

$$\begin{aligned} E_\mu[g_i] &\geqslant re^{-\lambda/2} \int_{c_i^2}^{\infty} \frac{v^{-1}}{2^{3/2}\Gamma(3/2)} e^{-v/2}\, dv \\ &\quad + r \cdot (\lambda/2) \sum_{k=1}^{\infty} \frac{(\lambda/2)^{k-1}}{(k-1)!} e^{-\lambda/2} b_k \bar{G}_{2k}(c_i^2) \\ &= I_1 + I_2 \qquad \text{(say)}, \end{aligned}$$

where

$$b_k = \frac{\Gamma(k)}{2^{3/2} k \Gamma(3/2+k)}.$$

Here

$$I_1 \geqslant re^{-\lambda/2} \int_{c_i^2}^{1} \frac{v^{-1}}{2^{3/2}\Gamma(3/2)} e^{-1/2}\, dv = \frac{1}{\sqrt{2\pi e}} re^{-\lambda/2} \log \frac{1}{c_i^2}.$$

Also noting that

$$b_k \geqslant \tilde{b}_k = \frac{1}{2^{3/2} k^2 (k+1)}$$

for $k \geqslant 1$, and by putting

$$N = \min_{k \geqslant 1} (3/2)^{k-1} \tilde{b}_k < \infty,$$

we have

$$I_2 \geqslant \frac{N}{2} r\lambda \sum_{k=1}^{\infty} \frac{(\lambda/2)^{k-1}}{(k-1)!} \cdot \frac{1}{(3/2)^{k-1}} e^{-\lambda/2} \bar{G}_2(c_i^2) = \frac{N}{2} r\bar{G}_2(c_i^2)\, \lambda e^{-\lambda/6},$$

and hence

$$E_\mu[g_i] \geqslant \frac{1}{\sqrt{2\pi e}} r \log \frac{1}{c_i^2} \cdot e^{-\lambda/2} + \frac{N}{2} r\bar{G}_2(c_i^2)\, \lambda e^{-\lambda/6}. \tag{57}$$

By (54) and (57), we see

$$E_\mu[g_i - h_i] \geqslant \frac{1}{\sqrt{2\pi e}} r \log \frac{1}{c_i^2} \cdot e^{-\lambda/2} + \frac{N}{2} r\bar{G}_2(c_i^2)\, \lambda e^{-\lambda/6} - c_i F(c_i)\, e^{-\lambda/4},$$

which is, for a sufficiently small $r > 0$, positive for all λ.

The proof is completed.

Proof of (b). Let

$$d_1 = \max(r, \sqrt{p-2}), \qquad d_2 = \max\left(\frac{p-1}{p-2} r, \sqrt{p-2}\right).$$

Since

$$\begin{aligned} -\Delta R_{JS+} & \\ &\leqslant E_\mu\left[\chi_{\{\|x\| > d_i\}} \frac{(p-2)^2}{\|x\|^2} + \chi_{\{\|x\| \leqslant d_i\}} \max\left\{\frac{(p-2)^2}{\|x\|^2}, 2p - \|x\|^2\right\}\right] \\ &\leqslant E_\mu\left[\chi_{\{\|x\| > d_i\}} \frac{(p-2)^2}{\|x\|^2} + \chi_{\{\|x\| \leqslant d_i\}} \frac{p^2}{\|x\|^2}\right] \end{aligned}$$

for $i=1, 2$, and

$$-\Delta R_r \geqslant E_\mu\left[\chi_{\{\|x\|>d_1\}}\left\{\frac{(p-2)^2}{\|x\|^2}+\frac{r}{\|x\|^3} f_1\left(\frac{r}{\|x\|}\right)\right\}\right],$$

$$-\Delta R_r^\dagger \geqslant E_\mu\left[\chi_{\{\|x\|>d_2\}}\left\{\frac{(p-2)^2}{\|x\|^2}+\frac{r}{\|x\|^3} f_2\left(\frac{r}{\|x\|}\right)\right\}\right],$$

we see

$$\Delta R_{JS+}-\Delta R_r \geqslant E_\mu[\tilde{g}_1(\|x\|)-\tilde{h}_1(\|x\|)],$$

$$\Delta R_{JS+}-\Delta R_r^\dagger \geqslant E_\mu[\tilde{g}_2(\|x\|)-\tilde{h}_2(\|x\|)],$$

where

$$\tilde{g}_i(\|x\|)=\chi_{\{\|x\|>d_i\}} \frac{r}{\|x\|^3} f_i\left(\frac{r}{\|x\|}\right),$$

$$\tilde{h}_i(\|x\|)=\chi_{\{\|x\|\leqslant d_i\}} \frac{p^2}{\|x\|^2}, \qquad i=1, 2.$$

In the same manner as in the proof of (a), we have

$$\begin{aligned}
E_\mu[\tilde{g}_i-\tilde{h}_i] &\geqslant Mr\bar{G}_1(d_i^2)\, e^{-\lambda/6}-p^2 d_i^{p-2} F(d_i)\, e^{-\lambda/4} \\
&\qquad \text{for} \quad p\geqslant 4, \\
&\geqslant \frac{1}{\sqrt{2\pi e}}\, r \log \frac{1}{d_i^2}\cdot e^{-\lambda/2}+\frac{N}{2}\, r\bar{G}_2(d_i^2)\, \lambda e^{-\lambda/6}-9 d_i F(d_i)\, e^{-\lambda/4} \\
&\qquad \text{for} \quad p=3,
\end{aligned}$$

which is, for any fixed $d_i>0$, positive for a sufficiently large λ. The proof is completed.

ACKNOWLEDGMENTS

A part of this work was done while the first author was staying at the University of Iowa. The authors are grateful to Richard L. Dykstra, the host professor, for his support. They also thank two referees for their helpful comments and suggestions. In particular Remark 3.5 is due to a comment from one referee.

REFERENCES

1. A. J. Baranchik, A family of minimax estimators of the mean of a multivariate normal distribution, *Ann. Math. Statist.* **41** (1970), 642–645.
2. R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, "Statistical Inference under Order Restrictions," Wiley, London, 1972.
3. H. Bergström, "Weak Convergence of Measures," Academic Press, New York, 1982.
4. M. E. Bock, Employing vague inequality information in the estimation of normal mean vectors (estimators that shrink to closed convex polyhedra), *in* "Statistical Decision Theory and Related Topics—III, Vol. 1" (S. S. Gupta and J. O. Berger, Eds.), pp. 169–193, Academic Press, New York, 1982.
5. M. E. Bock, Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions, *J. Multivariate Anal.* **17** (1985), 127–147.
6. Y.-T. Chang, Stein-type estimators for parameters restricted by linear inequalities, *Keio Sci. Tech. Rep.* **34** (1981), 83–95.
7. W. James and C. Stein, Estimation with quadratic loss, *in* "Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1" (J. Neyman, Ed.), pp. 361–379, Univ. of California Press, Berkeley, 1961.
8. G. G. Judge, T. A. Yancey, M. E. Bock, and R. Bohrer, The non-optimality of the inequality restricted estimator under squared error loss, *J. Econometrics* **25** (1984), 165–177.
9. T. Kubokawa, A unified approach to improving equivariant estimators, *Ann. Statist.* **22** (1994), 290–299.
10. T. Kubokawa, The Stein phenomenon in simultaneous estimation: A review, *in* "Applied Statistical Science III" (S. E. Ahmed, M. Ahsanullah, and B. K. Sinha, Eds.), pp. 143–173, NOVA, New York, 1998.
11. S. Kuriki, One-sided test for the equality of two covariance matrices, *Ann. Statist.* **21** (1993), 1379–1384.
12. S. Kuriki and A. Takemura, Some geometry of the cone of nonnegative definite matrices and weights of associated $\bar{\chi}^2$ distribution, *Ann. Inst. Statist. Math.* **52** (2000), 1–14.
13. D. Q. Naiman, Volumes of tubular neighborhoods of spherical polyhedra and statistical inference, *Ann. Statist.* **18** (1990), 685–716.
14. C. P. Robert, "The Bayesian Choice: A Decision-Theoretic Motivation," Springer-Verlag, New York, 1994.
15. T. Robertson, F. T. Wright, and R. L. Dykstra, "Order Restricted Statistical Inference," Wiley, Chichester, 1988.
16. A. L. Rukhin, Admissibility: Survey of a concept in progress, *Internat. Statist. Rev.* **63** (1995), 95–115.
17. R. Schneider, "Convex Bodies: The Brunn–Minkowski Theory," Cambridge Univ. Press, Cambridge, UK, 1993.
18. D. Sengupta and P. K. Sen, Shrinkage estimation in a restricted parameter space, *Sankhyā Ser. A* **53** (1991), 389–411.
19. A. Shapiro, Towards a unified theory of inequality constrained testing in multivariate analysis, *Internat. Statist. Rev.* **56** (1988), 49–62.
20. Y. Sheena, Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix, *J. Japan Statist. Soc.* **25** (1995), 35–48.
21. C. M. Stein, Estimation of the mean of a multivariate normal distribution, *Ann. Statist.* **9** (1981), 1135–1151.
22. A. Takemura and S. Kuriki, Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives, *Ann. Statist.* **25** (1997), 2368–2387.
23. R. Webster, "Convexity," Oxford Univ. Press, Oxford, 1994.
24. H. Weyl, On the volume of tubes, *Amer. J. Math.* **61** (1939), 461–472.